

Marek Nahotko

marek.nahotko@uj.edu.pl

Instytut Informacji Naukowej i Bibliotekoznawstwa

Uniwersytetu Jagiellońskiego

<http://dx.doi.org/10.18778/8142-336-6.06>

DANE W CHMURZE CZY DANE ZLINKOWANE?

Abstract: Two phenomena related to the functioning of libraries in a wide area network were compared: cloud library and Linked Open Data. Both of these phenomena are intertwined, but they are different. They are linked by the tendency to move ever-increasing amounts of data into the Web environment. A new way to include the metadata to the cloud is Linked Open Data. It is about putting all the metadata in the OPEN network: bibliographic descriptions, authority files, dictionaries of all kinds (for example subject headings vocabularies). These data, after appropriate processing (RDF syntactic), form a global metadata cloud, which greatly changes how the Internet functions as an information environment, including the role of libraries.

Słowa kluczowe: chmura danych, semantyczny web, Linked Open Data, outsourcing biblioteczny, współdziałanie metadanych.

Wstęp

Truizmem jest twierdzenie, że biblioteki, podobnie jak znakomita większość sfer życia we współczesnym świecie, uzależnione są od nowych technologii. Mało tego, wielokrotnie twierdziłem, że biblioteki często były w awangardzie stosowania innowacji technologicznych, posiadając ku temu warunki ze względu na gromadzone masowo informacje i metadane. Dlatego właśnie pierwsze systemy komputerowe pojawiły się w bibliotekach niemal zaraz po tym, gdy stały się one zdolne do w miarę bezawaryjnej pracy, czyli od lat sześćdziesiątych XX w. Biblioteki utrzymywały hegemonię, jeśli nie monopol, w organizacji informacji naukowej.

Sytuacja zaczęła ulegać zmianie wraz z rozwojem technologii sieciowych, w szczególności systemów i aplikacji Web 2.0. Wówczas chyba po raz pierwszy bibliotekarze poculi, że aplikacje te, do spółki z wyszukiwarkami globalnymi, mogą zakończyć hegemonię bibliotek w dziedzinie organizacji informacji. Użytkownicy bibliotek zaczęli się przenosić do innych dostawców informacji, a biblioteki rozpoczęły nerwowo modyfikować interfejsy

swoich systemów, dostosowując je, z lepszym lub gorszym skutkiem, do zasad funkcjonujących w nowym środowisku informacyjnym (na przykład dodając wizerunki okładek, tagowanie, komentowanie).

Aby zrozumieć kierunki pożądanych zmian w bibliotece, należy uświadomić sobie, że w otaczającym środowisku również zachodzą zmiany, wpływające na potrzeby użytkowników i możliwości, jakimi dysponują. Jak twierdzą Alison Head i Michael Eisenberg, studenci-użytkownicy bibliotek wyszukiwanie w Internecie uważają za wygodne, stosunkowo proste i łatwo dostępne. Z drugiej strony, korzystając z wyszukiwarek internetowych, narzekają na zbyt wielkie zasoby (przeładowanie informacją), wręcz zalewające poszukującego¹. Być może jednak rację mają ci, którzy twierdzą, jak Paul Levinson, że nie przeładowanie, a swoiście rozumiane niedoładowanie informacyjne (information underload) jest prawdziwym problemem². Polega ono na tym, że odbiorcom brak wiedzy, by skutecznie i najefektywniej korzystać z bogactwa sieci. Wykazują oni na przykład brak umiejętności właściwej selekcji materiałów ze względu na ich jakość. Szczególne problemy stwarza odszukanie potrzebnej informacji, o istnieniu której wiadomo i która potrzebna jest natychmiast. Jak twierdzi Levinson, umysł ludzki jest jednak z natury dostosowany do rozwiązywania takich problemów – nasze mózgi pozwalają dostrzegać nam sens w oszałamiającym i straszliwym zamęciu świata, w którym żyjemy. Potrzebne są tylko odpowiednie narzędzia i umiejętność posługiwania się nimi.

Interfejsy użytkownika, jako część oprogramowania bibliotecznego służąca komunikacji systemu z użytkownikiem, stanowią o jego sposobach pracy, oferowanych przez system. Są to lokalnie instalowane aplikacje, od zawsze stanowiące część każdego zintegrowanego systemu bibliotecznego. Coraz częściej oferują one funkcje nieobecne w tradycyjnym OPAC, takie jak wyszukiwanie według rozszerzonych słów kluczowych (różnych form danego terminu, na przykład synonimicznych i bliskoznaczących), upodabniające wyszukiwanie do prowadzonego przy pomocy języka naturalnego, które jest jednak wspierane przez automatyczną kontrolę relacji semantycznych, zachodzących między terminami wyszukiwawczymi. Wykorzystuje się także szczegółowe dane, zawarte w rekordach MARC, do tworzenia faset z zawartości podpól oraz do stosowania określników jako terminów wyszukiwawczych.

¹ A. Head; M. Eisenberg, *What today's college students say about conducting research in the digital age*, Washington 2009, s. 2.

² P. Levinson, *Nowe nowe media*, Kraków 2010, s. 194.

Biblioteka w chmurze?

Przetwarzanie w chmurze nie jest nową technologią, która pojawiła się ostatnio w Web, raczej należałoby je uznać za nową formę przetwarzania³. Jest to rodzaj technologii przetwarzania komputerowego, która pozwala na współużytkowanie zasobów i usług w Internecie zamiast przetrzymywania ich na lokalnych serwerach bądź węzłach sieci lub komputerach osobistych. Takie zdalne połączenie serwerów, sieci, aplikacji i zasobów nazywane jest chmurą. Działa ona jako technologia gromadzenia zasobów służących udostępnianiu wielkich zasobów usług obliczeniowych i informacji na życzenie użytkowników. Podobnie chmurę obliczeniową definiuje amerykański National Institute of Standards and Technology (NIST), określając ją jako model umożliwiający wygodny, na żądanie, sieciowy dostęp do rozproszonego połączenia konfigurowalnych zasobów przetwarzania (sieci, serwerów, pamięci, aplikacji i usług), które mogą być natychmiast dostarczone i wykorzystane z minimalnym wysiłkiem przeznaczonym na zarządzanie lub interakcję z dostawcą usług⁴.

Aplikacje chmurowe są jednym z typowych zjawisk charakteryzujących wspomniany już Web 2.0. Używając takich popularnych serwisów Web 2.0, jak Google, Wikipedia, Twitter lub Facebook nabywamy umiejętności obcowania z chmurą danych, gdyż aplikacje te oraz olbrzymie bazy danych, które obsługują, umieszczane są w wielkich centrach przetwarzania danych, stanowiących sedno przetwarzania w chmurze. Dzięki temu, używając Gmail mamy dostęp do swojej poczty z każdego miejsca i komputera przyłączonego do Sieci. Bibliotekarze, zarówno jako prywatni użytkownicy, jak i jako pracownicy instytucji, coraz częściej muszą dokonywać wyboru pomiędzy aplikacjami działającymi lokalnie oraz w technologiach chmurowych.

Dostęp do danych i oprogramowania w chmurze można porównać do innych zasobów, do których dostęp uzyskujemy poprzez różnego rodzaju sieci. Takimi zasobami są na przykład energia elektryczna lub woda (z wodociągu). Dane biblioteczne przeniesione do chmury stają się zasobem przeznaczonym dla określonej społeczności, do którego dostęp jest kupowany na zasadzie wskazań licznika, co oznacza, że płaci się tylko za wielkość zużytej. Zużycie to następuje w rozmiarze i czasie niezbędnym użytkownikowi. Jeżeli ruch w Sieci z jakiegoś powodu rośnie, to mechanizmy chmury bardzo szybko reagują,

³ A. Kaushik, A. Kumar, *Application of cloud computing in libraries*, "International Journal of Information Dissemination and Technology" 2013, vol. 3, no 4, s. 270.

⁴ NIST SP 800-145, *A NIST definition of cloud computing*, http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf, [dostęp: 20.04.2017].

przydzielając w tym miejscu dodatkowe zasoby. Możliwe są też rozwiązania hybrydowe, w których wrażliwe dane osobowe obsługiwane są lokalnie, a OPAC, dane i usługi z nim związane udostępniane są w chmurze.

Gorzej się sprawa przedstawia, gdy dostęp do zasobów z jakiegoś powodu zostaje zamknięty. Może tu chodzić o awarię techniczną lub o nieopłacenie abonamentu za dostęp. W takiej sytuacji bibliotekę korzystającą z własnego serwera porównać można do gospodarstwa korzystającego z własnego generatora prądu lub posiadającego własną studnię. Jak wiadomo, takie rozwiązania dają niezależność, jednak są znacznie bardziej kosztowne i bardzo utrudniają wprowadzenie jednolitych standardów. Biblioteki nigdy nie należały do instytucji nazbyt bogatych, więc możliwości oszczędności z pewnością trafiają do przekonania. Oszczędzać można na sprzęcie i usługach (wyspecjalizowanych kadrach), oferowanych przez dostawców serwisów chmurowych. Przy okazji uzyskuje się elastyczność dostępu, o której była już mowa. Inaczej bowiem niż w sieci elektrycznej lub wodociągowej, usługi chmurowe nie są związane z konkretnym miejscem. To tak, jakby można było korzystać z własnego abonamentu na energię elektryczną bez względu na to, gdzie z niej korzystamy (na przykład poza domem).

Przeniesienie części usług do chmury pozwala także ukryć koszty tworzenia oraz aplikacji nowych technologii. Każdy, kto miał do czynienia z komputeryzacją wie, że sprzęt, zwłaszcza sieciowy, starzeje się (głównie moralnie) po kilku latach. Podobnie też powstają coraz nowsze techniki programowania, wymuszające tworzenie kolejnych wersji systemów bibliotecznych. Biblioteki, w ramach stałego abonamentu, problemy te przerzucają na dostawcę, czyli jakby rozkładają na raty. Dostawca także osiąga korzyści, gdyż z jego punktu widzenia lepiej mieć jedną wersję oprogramowania w chmurze, stale fachowo zarządzaną i aktualizowaną, niż dziesiątki a czasem setki pojedynczych instalacji, które z biegiem czasu różnicują się coraz bardziej. Rozwiązania chmurowe bardzo ułatwiają więc życie dostawcom technologii i usług dla bibliotek.

W bibliotekach wyróżnia się następujące obszary zastosowań przetwarzania w chmurze:

1. Tworzenie bibliotek cyfrowych i repozytoriów. Obiekty cyfrowe, zazwyczaj skany zbiorów bibliotecznych, przenoszone są do chmury wraz z metadanymi. W ten sposób działać może DSpace.
2. Wyszukiwanie w komputerowych katalogach bibliotecznych – lokalnych i centralnych. Można uznać, że w ten sposób działa katalog zbiorów polskich bibliotek naukowych Nukat. Jest to miejsce, w którym rozpowszechniane są zasoby i usługi wspólnie tworzone przez polskie biblioteki współuczestniczące.

3. Obsługa strony WWW jest jednym z najwcześniejszych zastosowań chmury. Tworzenie i obsługa strony, wraz z jej utrzymywaniem na serwerze jest zlecana firmie zewnętrznej.
4. Wyszukiwanie treści naukowych w serwisach baz danych lub w Google Scholar.
5. Przechowywanie plików w celu ich rozpowszechniania lub innym, na przykład archiwizacji. Najpopularniejszym serwisem tego typu jest Flickr. Jak już o tym była mowa, biblioteki na serwerach dostawców usług przechowują swoje zasoby, począwszy od plików OPAC, a skończywszy na zdjęciach na Facebooku. Wyspecjalizowane serwisy, takie jak LOCKSS lub Portico pozwalają na cyfrową archiwizację zasobów.
6. Biblioteki przy pomocy chmury mogą budować sieć powiązań pomiędzy bibliotekarzami lub innymi osobami zainteresowanymi, w tym użytkownikami. Stosowane mogą być narzędzia tworzenia sieci społecznych. Najbardziej znane to Facebook i Twitter.
7. Przeniesienie wszystkich prac i usług do chmury, takich jak gromadzenie, katalogowanie, udostępnianie zasobów cyfrowych. Możliwe jest także zorganizowanie współpracy z dostawcami zewnętrznymi w oparciu o przekazywanie danych w chmurze.

To ostatnie zastosowanie chmury w szczególności prowadzi do głębokich zmian organizacyjnych w bibliotece, co wiąże się z przekazywaniem wykonywania wielu prac i zadań biblioteki do realizacji na zewnątrz (outsourcing). Jest to sposób zarządzania polegający na wykorzystaniu zasobów zewnętrznych zamiast wewnętrznych, poprzez zlecenie wyspecjalizowanym podmiotom wykonania całości lub części określonych zadań, funkcji lub procesów⁵. Taki sposób zarządzania organizacją pozwala skupić siły na realizacji jej podstawowych działań i zadań, inne zlecając firmom bardziej wyspecjalizowanym. Tutaj powstaje problem wyznaczenia możliwie ścisłej granicy pomiędzy tym, co jest podstawowym, niezbywalnym obszarem działań, a co można uznać za prace pomocnicze w bibliotece. Artur Jazdon przyjął, że tymi pierwszymi są opracowanie rzeczowe i formalne, jednak są takie biblioteki, które kopiują z katalogów centralnych opisy praktycznie wszystkich gromadzonych nowości. Podobnie wiele bibliotek całkowicie przerzuciło prace związane z prenumeratą i udostępnianiem czasopism (co ułatwiło ich funkcjonowanie w wersji elektronicznej) na firmy zewnętrzne, funkcjonujące w chmurze. W takiej sytuacji biblioteka staje się pośrednikiem między funduszodawcą a dostawcą w przekazywaniu pieniędzy na prenumeratę. Obecnie coraz częściej przyjmuje się, że najbardziej charakterystyczne funkcje biblioteki wiążą

⁵ A. Jazdon, *Outsourcing w bibliotece*, „Biblioteka” 2006, nr 10 (19), s. 105.

się z pracą z czytelnikiem, świadczeniem najszerzej rozumianych usług informacyjnych. W takim razie realizacji tej funkcji powinno się podporządkować wszystkie inne zadania oraz nie wykonywać jej w outsourcingu i w miarę możliwości nie w chmurze, chyba że jest to chmura biblioteczna. Jak bowiem stwierdza Jazdon, jeśli usługi informacyjne będzie się powierzać podmiotom zewnętrznym, to najprawdopodobniej będą źle zrealizowane.

Opisane zjawiska powodują bardzo duże zmiany w działalności bibliotek. Niezmienne pozostaje to, że zasoby biblioteczne nadal są zamknięte, zarówno z konceptualnego, jak i technicznego punktu widzenia. Zostają tylko przeniesione z lokalnego serwera na inny, chmurowy serwer. Z konceptualnego punktu widzenia należy zgodzić się, że standardy biblioteczne, na podstawie których tworzy się między innymi zasoby metadanych, pochodzą z czasów, gdy zasoby te miały funkcjonować w zamknięciu i były dostosowywane do przetwarzania (zarówno tworzenia, jak i wykorzystania) przez człowieka. Takie katalogi pozwalają na dostęp tylko do cyfrowych reprezentacji katalogów kartkowych⁶. Przetwarzanie komputerowe miało najwyżej charakter pomocniczy. Standardy te, zaimplementowane w systemach komputerowych z wykorzystaniem MARC, spowodowały wzrost duplikacji metadanych, niespójność danych, brak właściwej szczegółowości i złożoności⁷. Warto zaznaczyć, że z technicznego punktu widzenia, stosowane standardy mają charakter zamknięty, stosowane są tylko w środowisku bibliotekarskim i z wielkim trudem dają się przenosić do nowych standardów, używanych powszechnie poza tym środowiskiem, w Web. Wynika z tego potrzeba otwarcia zasobów bibliotecznych na nowe standardy pracy w chmurze, zwane ogólnie Linked Open Data (LOD).

Biblioteka zlinkowana?

W środowisku bibliotecznym na świecie i w Polsce, w ostatnich latach coraz częściej mówi się o rozwiązaniach dotyczących wprowadzenia zasobów bibliotecznych, zarówno danych, jak i metadanych, do tak zwanej sieci semantycznej (Semantic Web, Web 3.0). Z punktu widzenia zasobów metadanych jest to ich otwarcie na swobodne przetwarzanie przez odejście od trybu dostępu

⁶ H. Moulaison, A. Million, *The disruptive qualities of Linked Data in the library environment: analysis and recommendations*, „Cataloging & Classification Quarterly” 2014, vol. 52, s. 367.

⁷ G. Alemu [et al.], *Linked Data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models*, World Library and Information Congress 78th IFLA General Conference, Helsinki 2012.

rekord po rekordzie w kierunku wykorzystania relacji między elementami metadanych. Związane jest to z zastosowaniem RDF (Resource Description Framework) i pracami dotyczącymi modelowania metadanych z wykorzystaniem tego standardu⁸. Dzięki uproszczeniu wielu technologii sieci semantycznej przez W3C⁹, możliwym stało się powstanie i rozwój idei Linked Open Data), praktyki stosowania identyfikatorów używanych w Web, modelowania przepływu danych w różnych dyscyplinach i zastosowaniach oraz używania ontologii do projektowania metadanych funkcjonujących na zasadzie globalnego grafu. Linked Data określane jest jako dane opublikowane według zasad przyjętych dla ułatwienia tworzenia połączeń pomiędzy zestawami danych, zbiorami elementów i słownikami wartości elementów.

Chociaż Linked Data dotyczy wszelkiego rodzaju danych (nie tylko bibliograficznych) publikowanych w Web, to ożywienie w ostatnich latach obserwowane w tworzeniu nowych narzędzi, słowników i zasobów danych idea ta zawdzięcza szerokiemu i aktywnemu włączeniu się do prac osób ze środowiska bibliotekarskiego, które wniosły do tej działalności rozwiązania stosowane w zakresie danych i metadanych bibliotecznych. To dla przedstawicieli tego środowiska w 2010 r., w ramach konsorcjum W3C, powstał projekt Library Linked Data Incubator Group. Ma on wspomagać rozwój współdziałania metadanych bibliotecznych w Sieci¹⁰.

Linked Data ma służyć tworzeniu danych dotyczących wszelkich obiektów, nie tylko bibliograficznych, ale także na przykład ludzi, organizacji, procesów i pojęć. Zasady Linked Data w zastosowaniu do danych bibliograficznych przedstawił Marcin Roszkowski¹¹. Są to dane utworzone i opublikowane w taki sposób, aby umożliwić tworzenie połączeń pomiędzy zestawami danych oraz słownikami. Te ostatnie rozumiane są szerzej niż to było dotąd przyjęte w bibliotekarstwie, gdyż nazwą tą obejmuje się zarówno słowniki stosowane na poziomie pragmatyki, zawierające wartości metadanych, zwane także słownikami kontrolowanymi (kartoteki haseł wzorcowych, słowniki języków informacyjno-wyszukiwawczych, wykazy kodów nazw języków, nazw geograficznych i inne), jak również zestawy pojęć, stosowane na poziomie semantycznym, dotąd nazywane formatami lub schematami (wykazy elementów metadanych, pól i podpól formatów metadanych). W ten sposób do chmury

⁸G. Dunsire [et al.], *Linked Data vocabulary management: infrastructure support, data integration, and interoperability*, „Information Standards Quarterly” 2012, vol. 24, no 2/3, s. 4.

⁹WWW Consortium to organizacja powołana do tworzenia standardów dla Web, <http://www.w3.org/>, [dostęp: 12.04.2017].

¹⁰M. Roszkowski, *Linked Data – model danych powiązanych w Semantic Web*, „Zagadnienia Informacji Naukowej” 2010, nr 2, s. 61.

¹¹M. Roszkowski, *op. cit.*, s. 53 i nast.

danych trafiają słowniki zarówno poziomu semantycznego, jak i pragmatycznego metadanych, kodowane przy użyciu języków kodowania, a więc dostępne do bezpośredniego przetwarzania komputerowego.

Każdy element słownika jest jednoznacznie identyfikowany przy pomocy identyfikatora; na ogół stosowany jest URI (Uniform Resource Identifier) ze względu na jego otwartość (może być przydzielany oddolnie, przez użytkownika¹², bez pośrednictwa żadnej organizacji). Dzięki identyfikacji zasoby mogą być swobodnie cytowane przez inne zasoby, przez co metadane są powszechniej dostępne. Relacje pomiędzy identyfikowanymi elementami (dla danych bibliograficznych: dzieła, osoby, instytucje, miejsca, elementy treści i temu podobne) opisywane są przy pomocy języków, takich jak RDF i OWL (Web Ontology Language), których wyrażenia mogą być przetwarzane bezpośrednio przez komputery. W komunikacji naukowej słowniki wszystkich rodzajów i procesy zarządzania nimi odgrywają zasadniczą rolę. Nadają one znaczenie zbiorom danych, przez co posiadają wielkie znaczenie dla zapisów wiedzy naukowej i dla kultury¹³. Słowniki jako całość są natomiast opisywane przy pomocy metadanych gromadzonych w projekcie Linked Open Vocabularies¹⁴ (LOV), który prowadzi portal umożliwiający wyszukiwanie około sześciuset słowników ustrukturyzowanych.

Według Thomasa Bakera i grupy współpracowników¹⁵, Linked Open Data ma wiele zalet w porównaniu z dotychczas stosowanymi i rozwijanymi od lat technologiami, służącymi tworzeniu i rozpowszechnianiu metadanych bibliotecznych, gdyż stanowi ich rozwinięcie z myślą o zastosowaniu w sieciach globalnych. Dane zlinkowane są łatwe do rozpowszechniania, rozszerzalne i możliwe do wielokrotnego stosowania w wielu aplikacjach. Przynajmniej częściowo niezależne są od stosowanego języka naturalnego, gdyż URI budowany może być z dowolnych symboli, także liczbowych. Prace nad tworzeniem opisów zasobów mogą być realizowane na zasadzie współpracy bibliotek oraz innych instytucji, a także użytkowników. Ten rozproszony model danych

¹² URI oczywiście mogą i są zarządzane również przez instytucje. Bardzo ważną cechą URI jest jego stabilność, dlatego też użytkownicy są zachęceni do rozsądnego tworzenia własnych przestrzeni nazw dla swoich URI, które powinny być niepowtarzalne i niezmiennie. Te właśnie cechy są łatwiej osiągalne pod nadzorem instytucjonalnym.

¹³ D. Lovins, D. Hillmann, *Broken-world vocabularies*, „D-Lib Magazine” 2017, vol. 23, no 3/4.

¹⁴ Linked Open Vocabularies, <http://lov.okfn.org/dataset/lov/>, [dostęp: 12.04.2017].

¹⁵ T. Baker [et al.], *Library Linked Data Incubator Group final report*, Library Linked Data Incubator Group, <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>, [dostęp: 28.05.2013].

pozwala każdemu mówić cokolwiek o czymkolwiek, co musi powodować zróżnicowanie opinii i brak zgodności co do znaczenia elementów¹⁶. Jest to sytuacja nietypowa dla środowiska bibliotekarskiego, gdzie rozwiązania reprezentują raczej bardziej obiektywistyczny punkt widzenia. We współczesnym, otwartym środowisku sieci globalnych coraz większą rolę odgrywa punkt widzenia związany z kierunkiem filozoficznym zwanym społecznym konstrukttywizmem, w którym znaczenia dla zjawisk i obiektów świata otaczającego są wyznaczane przez ludzi; nie istnieją niezależnie od nas, od naszej świadomości¹⁷. Wówczas współdziałanie metadanych zapewniane jest nie poprzez narzucanie standardów, lecz przez współpracę w różnorodności.

Twórcy metadanych w Linked Open Data mogą udostępniać je w małych porcjach, jako indywidualne stwierdzenia (wypowiedzi). Wypowiedzi te przyjmują formę trójek RDF, składających się z podmiotu (opisywany zasób, na przykład dzieło: jego identyfikator), predykatu lub własności (relacja, na przykład ma autora) i obiektu (jak nazwa lub identyfikator autora¹⁸). Zamiast wymiany całych rekordów, zawierających pełny opis, jak to ma miejsce obecnie, możliwe jest więc dostarczanie tylko pojedynczych stwierdzeń o zasobie (trójek RDF), a wszystkie stwierdzenia o zasobie, mogące pochodzić z wielu źródeł, są agregowane w globalnym grafie. Nowe systemy biblioteczne mogą funkcjonować w oparciu o stwierdzenia istniejące w chmurze, uzupełniane danymi przechowywanymi lokalnie. Dzięki gęstej sieci połączeń (linków) do danych komplementarnych, tworzonych przez wiarygodne źródła (biblioteki) wzrasta wartość danych ponad to, co oferować może prosta suma zasobów traktowanych indywidualnie (bez powiązań). Do danych o osobie jako autorze mogą być bowiem dołączane dane o niej jako pracowniku, rodzicu, hobbyście i wiele innych. Łatwo zauważyć, że bardzo zmienia to sposób funkcjonowania kartotek wzorcowych jako słowników LOD.

Z punktu widzenia współdziałania metadanych, tworzonych w tym środowisku, ważne jest oddzielenie semantyki od syntaktyki metadanych. W Linked Data znaczenie metadanych (semantyka) jest oddzielone od ich syntaktyki, dzięki czemu zmiana struktur nie powoduje modyfikacji znaczenia. Tylko pozornie różnica tych sytuacji polega na zamianie jednych standardów innymi.

¹⁶ D. Allemang, J. Hendler, *Semantic Web for working ontologist: effective modeling in RDFS and OWL*, Amsterdam 2008, s. 7.

¹⁷ T. Duffy, D. Jonassen, *Objectivist and constructivist conceptions of learning and instruction*, [w:] *Constructivism and the technology of instruction: a conversation*, red. T. Duffy, D. Jonassen, Hillsdale 1992, s. 3.

¹⁸ M. Nahotko, *Opis dokumentów elektronicznych. Teoretyczny model i możliwości jego aplikacji*, Kraków 2006, s. 27.

W rzeczywistości nowe standardy mają jedną, ale bardzo ważną nową cechę – zarówno semantyka, jak i pragmatyka kodowane są przy pomocy tej samej syntaktyki (głównie XML/RDF, choć istnieją inne serializacje), a co więcej, jest to syntaktyka, która umożliwia automatyczne przetwarzanie danych (dzięki jej rozumieniu przez komputery), bez udziału człowieka. Opasłe, drukowane tablice Uniwersalnej Klasyfikacji Dziesiętnej (UKD) przestają być potrzebne, powstaje ich wersja w Linked Data z użyciem SKOS (Simple Knowledge Organization System). Dawne schematy metadanych nadal są używane (łącznie ze strukturą pól i podpól MARC 21), W3C (The World Wide Web Consortium) nawet rekomenduje ich dalsze stosowanie zamiast tworzenie nowych od początku¹⁹. Innowacją jest to, że struktury te są kodowane w RDFS lub OWL (zmianie uległa syntaktyka), przez co mogą stać się ontologiami umieszczanymi w chmurze danych. Do każdego symbolu UKD i do każdej relacji wyrażonej strukturą MARC (a więc do etykiety każdego pola tego formatu) oraz między tymi symbolami i etykietami można poprowadzić kwalifikowany link, czyli taki, który nazywa relację występującą pomiędzy połączonymi elementami.

Tak zorganizowane środowisko metadanych umożliwia dwojakiego rodzaju współdziałanie metadanych. Pierwszy sposób na zapewnienie współdziałania związany jest z istnieniem słowników dwóch opisanych rodzajów, przydatnych szczególnie, gdy chcemy łączyć dane z różnych obszarów (dziedzin, zastosowań). Użytkownik WWW może przeglądać jego zasoby bez potrzeby znajomości technologii i wynikających z nich struktur, stanowiących podstawę działania Internetu. Podczas przeglądania swobodnie przekraczane są granice między zasobami, bez względu na ich fizyczne oddalenie. Na tej samej zasadzie można w Linked Data przeglądać zbiory danych podążając za linkami od jednego zasobu do innego, nawet gdy są fizycznie umieszczone w różnych miejscach i zapisane w różnych formatach²⁰. Zasoby znajdują się w chmurze, ale są otwarte.

Odkrywanie (discovery) jest podstawową zasadą LOD i głównym sposobem aktywności użytkownika. Polega ono na podążaniu za linkami od jednego źródła danych do innego. Działa analogicznie do pracy robotów wyszukiwarek podczas wyszukiwania nowych stron WWW do indeksowania. W informatologii stosuje się termin przeglądanie (browsing). Technika ta spełnia swoje

¹⁹ *Data on the Web best practices*, W3C Recommendation 2017, <https://www.w3.org/TR/2017/REC-dwbp-20170131/>, [dostęp: 12.04.2017].

²⁰ E. Bermes, *Convergence and interoperability: a Linked Data perspective*, [w:] *World Library and Information Congress: 77th IFLA General Conference and Assembly*, Puerto Rico, 13–18 August 2011, <http://conference.ifla.org/past/ifla77/149-bermes-en.pdf>, [dostęp: 28.05.2013].

zadania w rozległym środowisku silnie zlinkowanych stron realizujących potrzeby użytkowników, o co dba algorytm wyszukiwarki.

Tego rodzaju zasobami danych są także zawartości słowników obu wcześniej wymienionych rodzajów. Umożliwiają one współdziałanie dzięki temu, że działają jak centrum przełącznikowe, umieszczone w chmurze danych, łączące dane wyrażone zgodnie z różnymi semantykami danych. Takie centrum umożliwia w Linked Data nawigację od jednego zbioru metadanych do innego, przez podążanie za odnośnikami, czyli URI, nawet jeżeli dane połączone do centrum są zróżnicowane. Wyobraźmy sobie na przykład, że istnieją dwa zasoby, w których znajdują się między innymi symbole UKD. Bez względu na to, w jakich miejscach różnych struktur metadanych (polach, elementach) umieszczone zostały te symbole, odsyłają one przy pomocy URI do ontologii UKD, wykonanej w SKOS i dostępnej w Internecie. Dzięki temu, po pierwsze, następuje ujednolicenie znaczenia (semantyki) danych, po drugie, możliwe jest wyszukiwanie podobnych danych poprzez proste podążanie za odnośnikami URI. Sposób ten pozwala na unikanie niejednorodności dzięki porozumieniu ontologicznemu²¹.

Co więcej, jako centrum przełącznikowe w Linked Data, może funkcjonować dowolny zasób udostępniony dla użytkowników. Nie istnieje jedno, główne centrum, ale wiele takich centrów, połączonych ze sobą hiperlinkami. Podążanie od jednego centrum do innego wykorzystując linki łączące dane, które tam się znajdują, pozwala na odnajdywanie nowej informacji w sposób intuicyjny. Jest to druga metoda zapewnienia współdziałania metadanych, nazwana podążaniem za własnym nosem. Suma zasobów danych stosujących RDF i URI stanowi globalny graf informacyjny, który może być w sposób nieograniczony przeglądany przez użytkowników (ludzi) i automatyczne aplikacje podążające za odnośnikami URI. Do jednostki identyfikującej osobę można na przykład przyłączyć obiekty, miejsca, czy inne osoby przy pomocy relacji: bycia autorem, pracownikiem, ojcem lub matką, mieszkańcem, fanem i dowolnej innej.

Twórcy danych w Linked Data powinni przede wszystkim sprawdzić, czy potrzebne im terminy nie zostały już zdefiniowane w jednym z istniejących słowników; w takim przypadku należy je użyć do reprezentacji danych. Jeżeli odpowiednich terminów brak, to należy utworzyć własny słownik, którego terminy zostaną połączone z zasobami istniejących słowników. Gdyby w późniejszym czasie okazało się, że jakiś inny słownik zawiera te same terminy, które zostały umieszczone w naszym słowniku, należy utworzyć link RDF pomiędzy URI terminu w obu słownikach, ze wskazaniem na ekwiwalentność łączonych

²¹ T. Heath, Ch. Bizer, *Linked Data. Evolving the Web into a Global Data Space*, San Rafael, CA. 2011, s. 24.

terminów przy pomocy relacji przewidzianych w OWL, RDFS lub SKOS. Autorytarność takich prac zapewnić mają nowe zasady publikowania danych, polegające w szczególności na zaopatrywaniu trójek RDF w metadane kontekstowe, co przewidywane jest w projekcie tak zwanej nanopublikacji²².

Już dostosowanie słowników, używanych dotąd w bibliotekach, do zasad Linked Open Data i udostępnienie ich do powszechnego stosowania umożliwia współdziałanie na podstawie tak powstałych centrów przełącznikowych od razu, bez żadnych dodatkowych prac. Według zasad, stosowanych w Linked Data, jeżeli dwa zasoby danych, stosujących różne formaty, używają tego samego słownika wartości metadanych (jak w powyższym przykładzie – tablice UKD), to istnieje możliwość przechodzenia od jednego zbioru danych do innego poprzez linki od tych zasobów do tego słownika. Jedynym wymogiem jest stosowanie identyfikatorów URI.

Zakończenie

Jak wynika z przedstawionego tekstu, w działalności bibliotecznej i informacyjnej zbliżamy się do rozwiązań polegających na przenoszeniu zasobów informacyjnych – danych i metadanych oraz związanych z nimi usług, do sieci globalnych. Biblioteki dysponują zasobami, które po konwersji do obowiązujących standardów Web mogą być natychmiast przenoszone do LOD. Wzbogaca to globalne środowisko informacyjne, ale także bardzo istotnie zmienia sposób funkcjonowania bibliotek i innych ośrodków informacji.

Biblioteka z miejsca, w którym gromadzi i opracowuje się zbiory, przekształca się w placówkę, w której prowadzi się usługi informacyjne ściśle dopasowane do potrzeb i poziomu różnych grup użytkowników. Działalność taka stanowi jeden ze sposobów przeciwdziałania nierównościami w dostępie do informacji w powstającym społeczeństwie informacyjnym. Misja ta formułowana była od dawna, ale dopiero obecnie, dzięki odciążeniu biblioteki od działań powtarzalnych i przeniesieniu zasobów do zlinkowanej chmury, stała się ona realnie wykonalna.

Dużym zmianom ulega profesja bibliotekarzy. Z jednej strony, po raz kolejny okazało się, że w tym środowisku od wielu dziesięcioleci powstawały rozwiązania (w tym przypadku zasoby semantyki), obecnie gotowe do przeniesienia do nowych standardów i wzbogacenia w kierunku tworzenia z nich ontologii. Z drugiej strony, bibliotekarze na rzecz społeczności sieciowej tracą kolejne funkcje, które od zawsze wydawały się immanentnie przypisane do tego

²² S. Gradmann, *From containers to content to context. The changing role of libraries in eScience and eScholarship*, „Journal of Documentation” 2014, vol. 70, no 2, s. 275.

zawodu: katalogowanie staje się katalinkowaniem, dostępnym każdemu. Wszelkie zasoby biblioteczne, informacji, danych i metadanych przenoszone są do Sieci tworząc zasoby globalnej, otwartej biblioteki cyfrowej.

Bibliografia

- Alemu G. [et al.], *Linked Data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models*, World Library and Information Congress 78th IFLA General Conference, Helsinki 2012.
- Allemang D., Hendler J., *Semantic Web for working ontologist: effective modeling in RDFS and OWL*, Amsterdam 2008.
- Baker T. [et al.], *Library Linked Data Incubator Group final report 2011*, Library Linked Data Incubator Group, <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>, [dostęp: 28.05.2013].
- Bermes E., *Convergence and interoperability: a Linked Data perspective*, [w:] *World Library and Information Congress: 77th IFLA General Conference and Assembly*, Puerto Rico, 13–18 August 2011, <http://conference.ifla.org/past/ifla77/149-bermes-en.pdf>, [dostęp: 28.05.2013].
- Duffy T., Jonassen, D., *Objectivist and constructivist conceptions of learning and instruction*, [w:] *Constructivism and the technology of instruction: a conversation*, red. T. Duffy, D. Jonassen, Hillsdale 1992.
- Dunsire G. [et al.], *Linked Data vocabulary management: infrastructure support, data integration, and interoperability*, „Information Standards Quarterly” 2012, vol. 24, no 2/3, s. 4–13.
- Gradmann S., *From containers to content to context. The changing role of libraries in eScience and eScholarship*, „Journal of Documentation” 2014, vol. 70, no 2, s. 241–260.
- Head A., Eisenberg M., *What today's college students say about conducting research in the digital age*, Washington 2009.
- Heath T., Bizer Ch., *Linked Data. Evolving the Web into a Global Data Space*, San Rafael, CA. 2011.
- Jazdon A., *Outsourcing w bibliotece*, „Biblioteka” 2006, nr 10 (19), s. 103–127.
- Kaushik A., Kumar A., *Application of cloud computing in libraries*, „International Journal of Information Dissemination and Technology” 2013, vol. 3, no 4, s. 270–273.
- Levinson P., *Nowe nowe media*, Kraków 2007.
- Lovins D., Hillmann D., *Broken-world vocabularies*, “D-Lib Magazine” 2017, vol. 23, no. 3/4.
- Moulaison H., Million A., *The disruptive qualities of Linked Data in the library environment: analysis and recommendations*, „Cataloging & Classification Quarterly” 2014, vol. 52, s. 367–387.

Nahotko M., *Opis dokumentów elektronicznych. Teoretyczny model i możliwości jego aplikacji*, Kraków 2006.

NIST SP 800-145, *A NIST definition of cloud computing*, [dostęp: 20.04.2017], http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf.

Roszkowski M., *Linked Data – model danych powiązanych w Semantic Web*, „Zagadnienia Informatyki Naukowej” 2010, nr 2, s. 52–68.